PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/12, C07K 16/18, 14/47, C12Q 1/68

A2

(11) International Publication Number: WO 99/38972

(43) International Publication Date: 5 August 1999 (05.08.99)

(21) International Application Number: PCT/US99/01619

(22) International Filing Date: 28 January 1999 (28.01.99)

(30) Priority Data:

60/072,910	28 January 1998 (28.01.98)	US
60/075,954	24 February 1998 (24.02.98)	US
60/080,114	31 March 1998 (31.03.98)	US
60/080,515	3 April 1998 (03.04.98)	US
60/080,666	3 April 1998 (03.04.98)	US

(71) Applicants *(for all designated States except US)*: CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US). HYSEQ INC. [US/US]; 675 Almanor Avenue, Sunnyvale, CA 94086 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: WILLIAMS, Lewis, T. [US/US]; 3 Miroflores, Tiburon, CA 94920 (US). ESCOBEDO, Jaime [CL/US]; 1470 Lavoma Road, Alamo, CA 94507 (US). INNIS, Michael, A. [US/US]; 315 Constance Place, Moraga, CA 94556 (US). GARCIA, Pablo, Dominguez [CL/US]; 882 Chenery Street, San Francisco, CA 94131 (US). SUDDUTH-KLINGER, Julie [US/US]; 280 Lexington Road, Kensington, CA 94707 (US). REINHARD, Christoph [DE/US]; 1633 Clinton Avenue, Alameda, CA 94501 (US). GIESE, Klause [DE/US]; 1009 Carolina Street, San Francisco, CA 94107 (US). RANDAZZO, Filippo [US/US]; 6363 Christie Avenue #2511, Emeryville, CA 94608 (US). KENNEDY, Giulia, C. [US/US]; 360 Castenada Avenue, San Francisco, CA 94116 (US). POT, David [CA/US]; 1565 5th Avenue #102, San Francisco, CA 94112 (US). KASSAM, Altaf [US/US]; 394 49th Street, Oakland, CA 94609 (US). LAMSON, George [US/US]; 232 Sandringham Drive, Moraga, CA 94556 (US). DRMANAC, Radoje [YU/US]; 850 East Greenwich Place, Palo Alto, CA 94303 (US). CRKVENJAKOV, Radomir [YU/US]; 762 Haverhill Drive, Sunnyvale, CA 94068 (US). DICKSON, Mark [US/US]; 1411 Gabilan Drive #B, Hollister, CA 95025 (US). DRMANAC, Snezana [YU/US]; 850 East Greenwich Place, Palo Alto, CA 94303 (US). LABAT, Ivan [YU/US]; 140 Acalanes Drive, Sunnyvale, CA 94086 (US). LESHKOWITZ, Dena [US/US]; 678 Durshire Way, Sunnyvale, CA 94087 (US). KITA, David [US/US]; 899 Bounty Drive, Foster City, CA 94404 (US). GARCIA, Veronica [ES/US]; 911 Shell Boulevard #102-0, Foster City, CA 96606 (US). JONES, William, Lee [US/US]; 4290 Albany Drive #P-146, San Jose, CA 95129 (US). STACHE-CRAIN, Birjit [DE/US]; 345 South Mary Avenue, Sunnyvale, CA 94086 (US).

(74) Agent: BLACKBURN, Robert, P.; Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: HUMAN GENES AND GENE EXPRESSION PRODUCTS II

(57) Abstract

This invention relates to novel human polynucleotides and variants thereof, their encoded polypeptides and variants thereof, to genes corresponding to these polynucleotides and to proteins expressed by the genes. The invention also relates to diagnostic and therapeutic agents employing such novel human polynucleotides, their corresponding genes or gene products, e.g., these genes and proteins, including probes, antisense constructs, and antibodies.

large numbers of oligodeoxyribonucleotides (oligos)) is described. The method is derived from DNA shuffling (Stemmer, *Nature* (1994) *370*:389-391), and does not rely on DNA ligase, but instead relies on DNA polymerase to build increasingly longer DNA fragments during the assembly process.

Appropriate polynucleotide constructs are purified using standard recombinant DNA techniques as described in, for example, Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual, 2nd Ed.*, (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY, and under current regulations described in United States Dept. of HHS, National Institute of Health (NIH) Guidelines for Recombinant DNA Research. The gene product encoded by a polynucleotide of the invention is expressed in any expression system, including, for example, bacterial, yeast, insect, amphibian and mammalian systems. Suitable vectors and host cells are described in U.S. Patent No. 5,654,173.

Bacteria. Expression systems in bacteria include those described in Chang *et al.*, *Nature* (1978) *275*:615; Goeddel *et al.*, *Nature* (1979) *281*:544; Goeddel *et al.*, *Nucleic Acids Res.* (1980) *8*:4057; EP 0 036,776; U.S. Patent No. 4,551,433; DeBoer *et al.*, *Proc. Natl. Acad. Sci. (USA)* (1983) *80*:21-25; and Siebenlist *et al.*, *Cell* (1980) *20*:269.

Yeast. Expression systems in yeast include those described in Hinnen *et al.*, *Proc. Natl. Acad. Sci. (USA)* (1978) *75*:1929; Ito *et al.*, *J. Bacteriol.* (1983) *153*:163; Kurtz *et al.*, *Mol. Cell. Biol.* (1986) *6*:142; Kunze *et al.*, *J. Basic Microbiol.* (1985) *25*:141; Gleeson *et al.*, *J. Gen. Microbiol.* (1986) *132*:3459; Roggenkamp *et al.*, *Mol. Gen. Genet.* (1986) *202*:302; Das *et al.*, *J. Bacteriol.* (1984) *158*:1165; De Louvencourt *et al.*, *J. Bacteriol.* (1983) *154*:737; Van den Berg *et al.*, *Bio/Technology* (1990) *8*:135; Kunze *et al.*, *J. Basic Microbiol.* (1985) *25*:141; Cregg *et al.*, *Mol. Cell. Biol.* (1985) *5*:3376; U.S. Patent Nos. 4,837,148 and 4,929,555; Beach and Nurse, *Nature* (1981) *300*:706; Davidow *et al.*, *Curr. Genet.* (1985) *10*:380; Gaillardin *et al.*, *Curr. Genet.* (1985) *10*:49; Ballance *et al.*, *Biochem. Biophys. Res. Commun.* (1983) *112*:284-289; Tilburn *et al.*, *Gene* (1983) *26*:205-221; Yelton *et al.*, *Proc. Natl. Acad. Sci. (USA)* (1984) *81*:1470-1474; Kelly and Hynes, *EMBO J.* (1985) *4*:475479; EP 0 244,234; and WO 91/00357.

Insect Cells. Expression of heterologous genes in insects is accomplished as described in U.S. Patent No. 4,745,051; Friesen *et al.*, "The Regulation of Baculovirus Gene Expression", in: *The Molecular Biology Of Baculoviruses* (1986) (W. Doerfler, ed.);

EP 0 127,839; EP 0 155,476; and Vlak *et al.*, *J. Gen. Virol.* (1988) *69*:765-776; Miller *et al.*, *Ann. Rev. Microbiol.* (1988) *42*:177; Carbonell *et al.*, *Gene* (1988) *73*:409; Maeda *et al.*, *Nature* (1985) *315*:592-594; Lebacq-Verheyden *et al.*, *Mol. Cell. Biol.* (1988) *8*:3129; Smith *et al.*, *Proc. Natl. Acad. Sci. (USA)* (1985) *82*:8844; Miyajima *et al.*, *Gene* (1987) *58*:273; and Martin *et al.*, *DNA* (1988) *7*:99. Numerous baculoviral strains and variants and corresponding permissive insect host cells from hosts are described in Luckow *et al.*, *Bio/Technology* (1988) *6*:47-55, Miller *et al.*, *Generic Engineering* (1986) *8*:277-279, and Maeda *et al.*, *Nature* (1985) *315*:592-594.

Mammalian Cells. Mammalian expression is accomplished as described in Dijkema *et al.*, *EMBO J.* (1985) *4*:761, Gorman *et al.*, *Proc. Natl. Acad. Sci. (USA)* (1982) *79*:6777, Boshart *et al.*, *Cell* (1985) *41*:521 and U.S. Patent No. 4,399,216. Other features of mammalian expression are facilitated as described in Ham and Wallace, *Meth. Enz.* (1979) *58*:44, Barnes and Sato, *Anal. Biochem.* (1980) *102*:255, U.S. Patent Nos. 4,767,704, 4,657,866, 4,927,762, 4,560,655, WO 90/103430, WO 87/00195, and U.S. RE 30,985.

Polynucleotide molecules comprising a polynucleotide sequence provided herein propagated by placing the molecule in a vector. Viral and non-viral vectors are used, including plasmids. The choice of plasmid will depend on the type of cell in which propagation is desired and the purpose of propagation. Certain vectors are useful for amplifying and making large amounts of the desired DNA sequence. Other vectors are suitable for expression in cells in culture. Still other vectors are suitable for transfer and expression in cells in a whole animal or person. The choice of appropriate vector is well within the skill of the art. Many such vectors are available commercially. The partial or full-length polynucleotide is inserted into a vector typically by means of DNA ligase attachment to a cleaved restriction enzyme site in the vector. Alternatively, the desired nucleotide sequence can be inserted by homologous recombination in vivo. Typically this is accomplished by attaching regions of homology to the vector on the flanks of the desired nucleotide sequence. Regions of homology are added by ligation of oligonucleotides, or by polymerase chain reaction using primers comprising both the region of homology and a portion of the desired nucleotide sequence, for example.

skill in the art, including those described in 5,242,974; 5,384,261; 5,405,783; 5,412,087; 5,424,186; 5,429,807; 5,436,327; 5,445,934; 5,472,672; 5,527,681; 5,529,756; 5,545,531; 5,554,501; 5,556,752; 5,561,071; 5,599,895; 5,624,711; 5,639,603; 5,658,734; WO 93/17126; WO 95/11995; WO 95/35505; EP 742287; and EP 799897. The arrays of the subject invention find use in a variety of applications, including gene expression analysis, drug screening, mutation analysis and the like, as disclosed in the above-listed exemplary patent documents.

In addition to the above nucleic acid libraries, analogous libraries of polypeptides are also provided, where the where the polypeptides of the library will represent at least a portion of the polypeptides encoded by "SEQ ID NOS:1-5252."

VII. Utilities

A. Use of Polynucleotide Probes in Mapping, and in Tissue Profiling

Polynucleotide probes, generally comprising at least 12 contiguous nucleotides of a polynucleotide as shown in the Sequence Listing, are used for a variety of purposes, such as chromosome mapping of the polynucleotide and detection of transcription levels. Additional disclosure about preferred regions of the disclosed polynucleotide sequences is found in the Examples. A probe that hybridizes specifically to a polynucleotide disclosed herein should provide a detection signal at least 5-, 10-, or 20-fold higher than the background hybridization provided with other unrelated sequences.

Probes in Detection of Expression Levels. Nucleotide probes are used to detect expression of a gene corresponding to the provided polynucleotide. In Northern blots, mRNA is separated electrophoretically and contacted with a probe. A probe is detected as hybridizing to an mRNA species of a particular size. The amount of hybridization is quantitated to determine relative amounts of expression, for example under a particular condition. Probes are used for in situ hybridization to cells to detect expression. Probes can also be used *in vivo* for diagnostic detection of hybridizing sequences. Probes are typically labeled with a radioactive isotope. Other types of detectable labels can be used such as chromophores, fluors, and enzymes. Other examples of nucleotide hybridization assays are described in WO92/02526 and U.S. Patent No. 5,124,246.

et al., *Methods in Molecular Biology* (1997) *68*:1, Boultwood, ed., Human Press, Totowa, NJ.

Polynucleotides are mapped to particular chromosomes using, for example, radiation hybrids or chromosome-specific hybrid panels. See Leach *et al.*, *Advances in Genetics*, (1995) *33*:63-99; Walter *et al.*, *Nature Genetics* (1994) *7*:22; Walter and Goodfellow, *Trends in Genetics* (1992) *9*:352. Panels for radiation hybrid mapping are available from Research Genetics, Inc., Huntsville, Alabama, USA. Databases for markers using various panels are available via the world wide web at http:/F/shgc-www.stanford.edu; and http://www-genome.wi.mit.edu/cgi-bin/contig/rhmapper.pl. The statistical program RHMAP can be used to construct a map based on the data from radiation hybridization with a measure of the relative likelihood of one order versus another. RHMAP is available via the world wide web at http://www.sph.umich.edu/group/statgen/software.

In addition, commercial programs are available for identifying regions of chromosomes commonly associated with disease, such as cancer. Polynucleotides based on the polynucleotides of the invention can be used to probe these regions. For example, if through profile searching a provided polynucleotide is identified as corresponding to a gene encoding a kinase, its ability to bind to a cancer-related chromosomal region will suggest its role as a kinase in one or more stages of tumor cell development/growth. Although some experimentation would be required to elucidate the role, the polynucleotide constitutes a new material for isolating a specific protein that has potential for developing a cancer diagnostic or therapeutic.

Tissue Typing or Profiling. Expression of specific mRNA corresponding to the provided polynucleotides can vary in different cell types and can be tissue-specific. This variation of mRNA levels in different cell types can be exploited with nucleic acid probe assays to determine tissue types. For example, PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes substantially identical or complementary to polynucleotides listed in the Sequence Listing can determine the presence or absence of the corresponding cDNA or mRNA.

For example, a metastatic lesion is identified by its developmental organ or tissue source by identifying the expression of a particular marker of that organ or tissue. If a

polynucleotide is expressed only in a specific tissue type, and a metastatic lesion is found to express that polynucleotide, then the developmental source of the lesion has been identified. Expression of a particular polynucleotide is assayed by detection of either the corresponding mRNA or the protein product. Immunological methods, such as antibody staining, are used to detect a particular protein product. Hybridization methods can be used to detect particular mRNA species, including but not limited to in situ hybridization and Northern blotting.

Use of Polymorphisms. A polynucleotide of the invention will be useful in forensics, genetic analysis, mapping, and diagnostic applications if the corresponding region of a gene is polymorphic in the human population. Particular polymorphic forms of the provided polynucleotides can be used to either identify a sample as deriving from a suspect or rule out the possibility that the sample derives from the suspect. Any means for detecting a polymorphism in a gene are used, including but not limited to electrophoresis of protein polymorphic variants, differential sensitivity to restriction enzyme cleavage, and hybridization to allele-specific probes.

B. Antibody Production

Expression products of a polynucleotide of the invention, the corresponding mRNA or cDNA, or the corresponding complete gene are prepared and used for raising antibodies for experimental, diagnostic, and therapeutic purposes. For polynucleotides to which a corresponding gene has not been assigned, this provides an additional method of identifying the corresponding gene. The polynucleotide or related cDNA is expressed as described above, and antibodies are prepared. These antibodies are specific to an epitope on the polypeptide encoded by the polynucleotide, and can precipitate or bind to the corresponding native protein in a cell or tissue preparation or in a cell-free extract of an in vitro expression system.

Immunogens for raising antibodies are prepared by mixing the polypeptides encoded by the polynucleotides of the present invention with adjuvants. Alternatively, polypeptides are made as fusion proteins to larger immunogenic proteins. Polypeptides are also covalently linked to other larger immunogenic proteins, such as keyhole limpet hemocyanin. Immunogens are typically administered intradermally, subcutaneously, or intramuscularly. Immunogens are administered to experimental animals such as rabbits,

In addition to the antibodies discussed above, genetically engineered antibody derivatives are made, such as single chain antibodies, according to methods well known in the art.

C. Use of Polynucleotides to Construct Arrays for Diagnostics

Polynucleotide arrays provide a high throughput technique that can assay a large number of polynucleotide sequences in a sample. This technology can be used as a diagnostic and as a tool to test for differential expression to determine function of an encoded protein. Arrays can be created by spotting polynucleotide probes onto a substrate (*e.g.*, glass, nitrocelllose, *etc.*) in a two-dimensional matrix or array having bound probes. The probes can be bound to the substrate by either covalent bonds or by non-specific interactions, such as hydrophobic interactions. Samples of polynucleotides can be detectably labeled (*e.g.*, using radioactive or fluorescent labels) and then hybridized to the probes. Double stranded polynucleotides, comprising the labeled sample polynucleotides bound to probe polynucleotides, can be detected once the unbound portion of the sample is washed away. Techniques for constructing arrays and methods of using these arrays are described in EP No. 0 799 897; PCT No. WO 97/29212; PCT No. WO 97/27317; EP No. 0 785 280; PCT No. WO 97/02357; U.S. Pat. No. 5,593,839; U.S. Pat. No. 5,578,832; EP No. 0 728 520; U.S. Pat. No. 5,599,695; EP No. 0 721 016; U.S. Pat. No. 5,556,752; PCT No. WO 95/22058; and U.S. Pat. No. 5,631,734.

As discussed in some detail above, arrays can be used to examine differential expression of genes and can be used to determine gene function. For example, arrays of the instant polynucleotide sequences can be used to determine if any of the provided polynucleotides are differentially expressed between a test cell and control cell (*e.g.*, cancer cells and normal cells). For example, high expression of a particular message in a cancer cell, which is not observed in a corresponding normal cell, can indicate a cancer specific protein. Exemplary uses of arrays are further described in, for example, Pappalarado *et al.*, *Sem. Radiation Oncol.* (1998) *8*:217; and Ramsay *Nature Biotechnol.* (1998) *16*:40.

D. Differential Expression

The polynucleotides of the invention can also be used to detect differences in expression levels between two cells, *e.g.*, as a method to identify abnormal or diseased tissue in a human. For polynucleotides corresponding to profiles of protein families, the

to, amniotic fluid, chorionic villi, blood, and the blastomere of an in vitro-fertilized embryo. The comparable normal polynucleotide-related gene is obtained from any tissue. The mRNA or protein is obtained from a normal tissue of a human in which the polynucleotide-related gene is expressed. Differences such as alterations in the nucleotide sequence or size of the same product of the fetal polynucleotide-related gene or mRNA, or alterations in the molecular weight, amino acid sequence, or relative abundance of fetal protein, can indicate a germline mutation in the polynucleotide-related gene of the fetus, which indicates a genetic predisposition to disease. Particular diagnostic and prognostic uses of the disclosed polynucleotides are described in more detail below.

E. Diagnostic, Prognostic, and Other Uses Based On Differential Expression

In general, diagnostic methods of the invention for involve detection of a level or amount of a gene product, particularly a differentially expressed gene product, in a test sample obtained from a patient suspected of having or being susceptible to a disease (*e.g.*, breast cancer, lung cancer, colon cancer and/or metastatic forms thereof), and comparing the detected levels to those levels found in normal cells (*e.g.*, cells substantially unaffected by cancer) and/or other control cells (*e.g.*, to differentiate a cancerous cell from a cell affected by dysplasia). Furthermore, the severity of the disease can be assessed by comparing the detected levels of a differentially expressed gene product with those levels detected in samples representing the levels of differentially gene product associated with varying degrees of severity of disease.

The term "differentially expressed gene" is intended to encompass a polynucleotide that can, for example, include an open reading frame encoding a gene product (*e.g.*, a polypeptide), and/or introns of such genes and adjacent 5' and 3' non-coding nucleotide sequences involved in the regulation of expression, up to about 20 kb beyond the coding region, but possibly further in either direction. The gene can be introduced into an appropriate vector for extrachromosomal maintenance or for integration into a host genome. In general, a difference in expression level associated with a decrease in expression level of at least about 25%, usually at least about 50% to 75%, more usually at least about 90% or more is indicative of a differentially expressed gene of interest, *i.e.*, a gene that is underexpressed or down-regulated in the test sample relative to a control sample. Furthermore, a difference in expression level associated with an increase in

expression of at least about 25%, usually at least about 50% to 75%, more usually at least about 90% and can be at least about 1 ½-fold, usually at least about 2-fold to about 10-fold, and can be about 100-fold to about 1,000-fold increase relative to a control sample is indicative of a differentially expressed gene of interest, *i.e.*, an overexpressed or up-regulated gene.

"Differentially expressed polynucleotide" as used herein means a nucleic acid molecule (RNA or DNA) having a sequence that represents a differentially expressed gene, *e.g.*, the differentially expressed polynucleotide comprises a sequence (*e.g.*, an open reading frame encoding a gene product) that uniquely identifies a differentially expressed gene so that detection of the differentially expressed polynucleotide in a sample is correlated with the presence of a differentially expressed gene in a sample. "Differentially expressed polynucleotides" is also meant to encompass fragments of the disclosed polynucleotides, *e.g.*, fragments retaining biological activity, as well as nucleic acids homologous, substantially similar, or substantially identical (*e.g.*, having about 90% sequence identity) to the disclosed polynucleotides.

Methods of the subject invention useful in diagnosis or prognosis typically involve comparison of the abundance of a selected differentially expressed gene product in a sample of interest with that of a control to determine any relative differences in the expression of the gene product, where the difference can be measured qualitatively and/or quantitatively. Quantitation can be accomplished, for example, by comparing the level of expression product detected in the sample with the amounts of product present in a standard curve. A comparison can be made visually; by using a technique such as densitometry, with or without computerized assistance; by preparing a representative library of cDNA clones of mRNA isolated from a test sample, sequencing the clones in the library to determine that number of cDNA clones corresponding to the same gene product, and analyzing the number of clones corresponding to that same gene product relative to the number of clones of the same gene product in a control sample; or by using an array to detect relative levels of hybridization to a selected sequence or set of sequences, and comparing the hybridization pattern to that of a control. The differences in expression are then correlated with the presence or absence of an abnormal expression pattern. A variety of different methods for determining the nucleic acid abundance in a sample are known to

Any of a variety of detectable labels can be used in connection with the various embodiments of the diagnostic methods of the invention. Suitable detectable labels include fluorochromes,(*e.g.* fluorescein isothiocyanate (FITC), rhodamine, Texas Red, phycoerythrin, allophycocyanin, 6-carboxyfluorescein (6-FAM), 2',7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein, 6-carboxy-X-rhodamine (ROX), 6-carboxy-2',4',7',4,7-hexachlorofluorescein (HEX), 5-carboxyfluorescein (5-FAM) or N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA)), radioactive labels, (*e.g.* ^{32}P, ^{35}S, ^{3}H, *etc.*), and the like. The detectable label can involve a two stage systems (*e.g.*, biotin-avidin, hapten-anti-hapten antibody, *etc.*)

Reagents specific for the polynucleotides and polypeptides of the invention, such as antibodies and nucleotide probes, can be supplied in a kit for detecting the presence of an expression product in a biological sample. The kit can also contain buffers or labeling components, as well as instructions for using the reagents to detect and quantify expression products in the biological sample. Exemplary embodiments of the diagnostic methods of the invention are described below in more detail.

Polypeptide detection in diagnosis. In one embodiment, the test sample is assayed for the level of a differentially expressed polypeptide. Diagnosis can be accomplished using any of a number of methods to determine the absence or presence or altered amounts of the differentially expressed polypeptide in the test sample. For example, detection can utilize staining of cells or histological sections with labeled antibodies, performed in accordance with conventional methods. Cells can be permeabilized to stain cytoplasmic molecules. In general, antibodies that specifically bind a differentially expressed polypeptide of the invention are added to a sample, and incubated for a period of time sufficient to allow binding to the epitope, usually at least about 10 minutes. The antibody can be detectably labeled for direct detection (*e.g.*, using radioisotopes, enzymes, fluorescers, chemiluminescers, and the like), or can be used in conjunction with a second stage antibody or reagent to detect binding (*e.g.*, biotin with horseradish peroxidase-conjugated avidin, a secondary antibody conjugated to a fluorescent compound, *e.g.* fluorescein, rhodamine, Texas red, *etc.*). The absence or presence of antibody binding can be determined by various methods, including flow cytometry of dissociated cells, microscopy, radiography, scintillation counting, *etc.* Any suitable alternative methods can

of qualitative or quantitative detection of levels or amounts of differentially expressed polypeptide can be used, for example ELISA, western blot, immunoprecipitation, radioimmunoassay, etc.

In general, the detected level of differentially expressed polypeptide in the test sample is compared to a level of the differentially expressed gene product in a reference or control sample, *e.g.*, in a normal cell (negative control) or in a cell having a known disease state (positive control).

<u>mRNA detection.</u> The diagnostic methods of the invention can also or alternatively involve detection of mRNA encoded by a gene corresponding to a differentially expressed polynucleotides of the invention. Any suitable qualitative or quantitative methods known in the art for detecting specific mRNAs can be used. mRNA can be detected by, for example, *in situ* hybridization in tissue sections, by reverse transcriptase-PCR, or in Northern blots containing poly A+ mRNA. One of skill in the art can readily use these methods to determine differences in the size or amount of mRNA transcripts between two samples. For example, the level of mRNA of the invention in a tissue sample suspected of being cancerous or dysplastic is compared with the expression of the mRNA in a reference sample, *e.g.*, a positive or negative control sample (*e.g.*, normal tissue, cancerous tissue, *etc.*).

Any suitable method for detecting and comparing mRNA expression levels in a sample can be used in connection with the diagnostic methods of the invention (see, *e.g.*, U.S. 5,804,382). For example, mRNA expression levels in a sample can be determined by generation of a library of expressed sequence tags (ESTs) from the sample, where the EST library is representative of sequences present in the sample (Adams, et al., (1991) *Science* 252:1651). Enumeration of the relative representation of ESTs within the library can be used to approximate the relative representation of the gene transcript within the starting sample. The results of EST analysis of a test sample can then be compared to EST analysis of a reference sample to determine the relative expression levels of a selected polynucleotide, particularly a polynucleotide corresponding to one or more of the differentially expressed genes described herein.

Alternatively, gene expression in a test sample can be performed using serial analysis of gene expression (SAGE) methodology (Velculescu et al., *Science* (1995)

Use of a single gene in diagnostic applications. The diagnostic methods of the invention can focus on the expression of a single differentially expressed gene. For example, the diagnostic method can involve detecting a differentially expressed gene, or a polymorphism of such a gene (*e.g.*, a polymorphism in an coding region or control region), that is associated with disease. Disease-associated polymorphisms can include deletion or truncation of the gene, mutations that alter expression level and/or affect activity of the encoded protein, *etc.*

Changes in the promoter or enhancer sequence that affect expression levels of an differentially gene can be compared to expression levels of the normal allele by various methods known in the art. Methods for determining promoter or enhancer strength include quantitation of the expressed natural protein; insertion of the variant control element into a vector with a reporter gene such as β-galactosidase, luciferase, chloramphenicol acetyltransferase, *etc.* that provides for convenient quantitation; and the like.

A number of methods are available for analyzing nucleic acids for the presence of a specific sequence, *e.g.* a disease associated polymorphism. Where large amounts of DNA are available, genomic DNA is used directly. Alternatively, the region of interest is cloned into a suitable vector and grown in sufficient quantity for analysis. Cells that express a differentially expressed gene can be used as a source of mRNA, which can be assayed directly or reverse transcribed into cDNA for analysis. The nucleic acid can be amplified by conventional techniques, such as the polymerase chain reaction (PCR), to provide sufficient amounts for analysis, and a detectable label can be included in the amplification reaction (*e.g.*, using a detectably labeled primer or detectably labeled oligonucleotides) to facilitate detection. The use of the polymerase chain reaction is described in Saiki, *et al.*, *Science* (1985) *239*:487, and a review of techniques can be found in Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual*, (1989) pp. 14.2. Alternatively, various methods are known in the art that utilize oligonucleotide ligation as a means of detecting polymorphisms, for examples see Riley *et al.*, *Nucl. Acids Res.* (1990) *18*:2887; and Delahunty *et al.*, *Am. J. Hum. Genet.* (1996) *58*:1239.

The sample nucleic acid, *e.g.* amplified or cloned fragment, is analyzed by one of a number of methods known in the art. The nucleic acid can be sequenced by dideoxy or other methods, and the sequence of bases compared to a selected sequence, *e.g.*, to a wild-

breast tissue, to discriminate between breast cancers with different cells of origin, to discriminate between breast cancers with different potential metastatic rates, etc.

G. Use of Polynucleotides to Screen for Peptide Analogs and Antagonists

Polypeptides encoded by the instant polynucleotides and corresponding full length genes can be used to screen peptide libraries to identify binding partners, such as receptors, from among the encoded polypeptides.

A library of peptides can be synthesized following the methods disclosed in U.S. Pat. No. 5,010,175 ('175), and in WO 91/17823. As described below in brief, one prepares a mixture of peptides, which is then screened to identify the peptides exhibiting the desired signal transduction and receptor binding activity. In the '175 method, a suitable peptide synthesis support (*e.g.*, a resin) is coupled to a mixture of appropriately protected, activated amino acids. The concentration of each amino acid in the reaction mixture is balanced or adjusted in inverse proportion to its coupling reaction rate so that the product is an equimolar mixture of amino acids coupled to the starting resin. The bound amino acids are then deprotected, and reacted with another balanced amino acid mixture to form an equimolar mixture of all possible dipeptides. This process is repeated until a mixture of peptides of the desired length (*e.g.*, hexamers) is formed. Note that one need not include all amino acids in each step: one can include only one or two amino acids in some steps (*e.g.*, where it is known that a particular amino acid is essential in a given position), thus reducing the complexity of the mixture. After the synthesis of the peptide library is completed, the mixture of peptides is screened for binding to the selected polypeptide. The peptides are then tested for their ability to inhibit or enhance activity. Peptides exhibiting the desired activity are then isolated and sequenced.

The method described in WO 91/17823 is similar. However, instead of reacting the synthesis resin with a mixture of activated amino acids, the resin is divided into twenty equal portions (or into a number of portions corresponding to the number of different amino acids to be added in that step), and each amino acid is coupled individually to its portion of resin. The resin portions are then combined, mixed, and again divided into a number of equal portions for reaction with the second amino acid. In this manner, each reaction can be easily driven to completion. Additionally, one can maintain separate "subpools" by treating portions in parallel, rather than combining all resins at each step.

This simplifies the process of determining which peptides are responsible for any observed receptor binding or signal transduction activity.

In such cases, the subpools containing, *e.g.*, 1-2,000 candidates each are exposed to one or more polypeptides of the invention. Each subpool that produces a positive result is then resynthesized as a group of smaller subpools (sub-subpools) containing, *e.g.*, 20-100 candidates, and reassayed. Positive sub-subpools can be resynthesized as individual compounds, and assayed finally to determine the peptides that exhibit a high binding constant. These peptides can be tested for their ability to inhibit or enhance the native activity. The methods described in WO 91/7823 and U.S. Patent No. 5,194,392 (herein incorporated by reference) enable the preparation of such pools and subpools by automated techniques in parallel, such that all synthesis and resynthesis can be performed in a matter of days.

Peptide agonists or antagonists are screened using any available method, such as signal transduction, antibody binding, receptor binding, mitogenic assays, chemotaxis assays, etc. The methods described herein are presently preferred. The assay conditions ideally should resemble the conditions under which the native activity is exhibited *in vivo*, that is, under physiologic pH, temperature, and ionic strength. Suitable agonists or antagonists will exhibit strong inhibition or enhancement of the native activity at concentrations that do not cause toxic side effects in the subject. Agonists or antagonists that compete for binding to the native polypeptide can require concentrations equal to or greater than the native concentration, while inhibitors capable of binding irreversibly to the polypeptide can be added in concentrations on the order of the native concentration.

The end results of such screening and experimentation will be at least one novel polypeptide binding partner, such as a receptor, encoded by a gene or a cDNA corresponding to a polynucleotide of the invention, and at least one peptide agonist or antagonist of the novel binding partner. Such agonists and antagonists can be used to modulate, enhance, or inhibit receptor function in cells to which the receptor is native, or in cells that possess the receptor as a result of genetic engineering. Further, if the novel receptor shares biologically important characteristics with a known receptor, information about agonist/antagonist binding can facilitate development of improved agonists/antagonists of the known receptor.

H. Pharmaceutical Compositions and Therapeutic Uses

Pharmaceutical compositions can comprise polypeptides, antibodies, or polynucleotides of the claimed invention. The pharmaceutical compositions will comprise a therapeutically effective amount of either polypeptides, antibodies, or polynucleotides of the claimed invention.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation is determined by routine experimentation and is within the judgment of the clinician. For purposes of the present invention, an effective dose will generally be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which can be administered without undue toxicity. Suitable carriers can be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in the art.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in *Remington's Pharmaceutical Sciences* (Mack Pub. Co., N.J. 1991).

SEQ ID NO:	Filing Date of Priority Appln	SEQ ID NO: in Priority Appln	Sequence Name	Clone Name	Cluster ID
2922	2/24/98	1195	RTA00000422F.e.23.1	M00001567D:B03	19246
2923	2/24/98	1168	RTA00000421F.b.06.1	M00001567A:B09	2113
2924	2/24/98	580	RTA00000403F.o.07.1	M00001579C:A01	39037
2925	2/24/98	531	RTA00000345F.n.12.1	M00001528A:C04	7337
2926	2/24/98	154	RTA00000340F.b.21.1	M00001533D:A08	8001
2927	2/24/98	19	RTA00000123A.k.23.1	M00001533A:G05	80313
2928	2/24/98	1265	RTA00000340F.d.07.1	M00001532D:A06	0
2929	2/24/98	1124	RTA00000123A.h.22.1	M00001532A:C01	17124
2930	2/24/98	1241	RTA00000408F.l.14.1	M00001530A:E10	12001
2931	2/24/98	534	RTA00000126A.g.7.1	M00001548A:H04	1902
2932	2/24/98	694	RTA00000418F.c.07.1	M00001529D:C05	73245
2933	2/24/98	1034	RTA00000124A.f.16.3	M00001536A:F11	47430
2934	2/24/98	790	RTA00000345F.n.08.1	M00001517A:B11	0
2935	2/24/98	613	RTA00000122A.j.22.1	M00001516A:F06	81151
2936	2/24/98	885	RTA00000122A.j.17.1	M00001516A:D02	62736
2937	2/24/98	1262	RTA00000122A.h.4.1	M00001514A:G03	33576
2938	2/24/98	135	RTA00000122A.d.5.1	M00001513A:F05	81155
2939	1/28/98	391	RTA00000179AF.c.20.3	M00001396A:C03	4009
2940	2/24/98	537	RTA00000408F.l.09.1	M00001530A:A09	75487
2941	2/24/98	683	RTA00000403F.j.21.1	M00001540D:E02	24723
2942	2/24/98	343	RTA00000422F.g.21.1	M00001583A:F07	17232
2943	2/24/98	226	RTA00000125A.k.10.1	M00001545A:F02	81644
2944	2/24/98	763	RTA00000135A.m.18.1	M00001545A:C03	19255
2945	2/24/98	156	RTA00000125A.k.1.1	M00001545A:B12	0
2946	2/24/98	597	RTA00000135A.l.1.2	M00001545A:B10	39426
2947	2/24/98	586	RTA00000125A.g.24.1	M00001544A:F05	80397
2948	2/24/98	467	RTA00000123A.n.13.2	M00001534A:D03	39167
2949	2/24/98	830	RTA00000347F.b.08.1	M00001541B:E05	17591
2950	2/24/98	997	RTA00000134A.l.9.1	M00001535A:D10	81814
2951	2/24/98	371	RTA00000403F.j.17.1	M00001539D:B10	38563
2952	2/24/98	33	RTA00000403F.j.15.1	M00001539B:G07	23840
2953	2/24/98	1209	RTA00000408F.n.05.2	M00001539A:H02	77883
2954	2/24/98	530	RTA00000408F.n.02.2	M00001539A:E01	76993
2955	2/24/98	1213	RTA00000135A.a.23.1	M00001537A:H05	27054
2956	2/24/98	347	RTA00000125A.n.4.1	M00001546A:D08	81984
2957	2/24/98	472	RTA00000135A.f.14.2	M00001542A:G12	79969
2958	2/24/98	243	RTA00000410F.c.14.1	M00001634A:H05	77809
2959	2/24/98	919	RTA00000410F.d.18.1	M00001635D:D05	75458
2960	2/24/98	825	RTA00000404F.k.22.2	M00001635D:C12	39084
2960	2/24/98	364	RTA00000404F.k.22.1	M00001635D:C12	39084
2961	2/24/98	825	RTA00000404F.k.22.2	M00001635D:C12	39084
2961	2/24/98	364	RTA00000404F.k.22.1	M00001635D:C12	39084
2962	2/24/98	595	RTA00000410F.d.10.1	M00001635B:H02	77561
2963	2/24/98	175	RTA00000410F.d.09.1	M00001635B:H01	76964
2964	2/24/98	206	RTA00000410F.b.15.1	M00001633C:F09	77100
2965	2/24/98	1083	RTA00000418F.j.20.1	M00001634D:D04	77101
2966	2/24/98	922	RTA00000410F.e.09.1	M00001636A:F08	76093
2967	2/24/98	1035	RTA00000404F.k.15.1	M00001634A:B04	18225
2968	2/24/98	1167	RTA00000410F.c.06.1	M00001633D:H06	77784
2969	2/24/98	53	RTA00000410F.c.04.1	M00001633D:G09	74099

	Nearest Neighbor (BlastN vs. Genbank)			Nearest Neighbor (BlastX vs. Non-Redundant Prot ins)		
SEQ ID	ACCESSION	DESCRIPTION	P VALUE	ACCESSION	DESCRIPTION	P VALUE
2917	X80283	P.polycephalum genomic DNA containing Taq I repetitive element	3.3	<NONE>	<NONE>	<NONE>
2918	U95094	Xenopus laevis XL-INCENP (XL-INCENP) mRNA, complete cds	6e-006	<NONE>	<NONE>	<NONE>
2919	U95094	Xenopus laevis XL-INCENP (XL-INCENP) mRNA, complete cds	6e-006	<NONE>	<NONE>	<NONE>
2920	Z97333	Homo sapiens RHCE gene	9e-020	113667	!!!! ALU CLASS B WARNING ENTRY !!!!	4e-005
2921	AF082350	Homo sapiens bone morphogenetic protein 15 precursor (BMP15) gene, exon 2 and complete cds	1	<NONE>	<NONE>	<NONE>
2922	L14684	Rattus norvegicus nuclear-encoded mitochondrial elongation factor G mRNA, complete cds.	0	585084	ELONGATION FACTOR G, MITOCHONDRIAL PRECURSOR (MEF-G) >gi|543383|pir||S40780 translation elongation factor G, mitochondrial - rat >gi|310102	9e-089
2923	D78335	Human mRNA for 5'-terminal region of UMK, complete cds	e-163	1718058	URIDINE KINASE (URIDINE MONOPHOSPHOKINASE) >gi|471981 (L31783) uridine kinase	7e-072
2924	U95094	Xenopus laevis XL-INCENP (XL-INCENP) mRNA, complete cds	2e-007	<NONE>	<NONE>	<NONE>

We Claim:

1. A library of polynucleotides, the library comprising the sequence information of at least one of SEQ ID NOS:1-3544, 3546-4510, 4512-4725, 4727-4748, and 4750-5252.

2. The library of claim 1, wherein the library is provided on a nucleic acid array.

3. The library of claim 1, wherein the library is provided in a computer-readable format.

4. The library of claim 1, wherein the library comprises a differentially expressed polynucleotide comprising a sequence selected from the group consisting of SEQ ID NOS:65, 174, 203, 252, 253, 387, 419, 420, 491, 552, 560, 581, 590, 648, 693, 726, 746, 990, 1095, 1124, 1205, 1354, 1387, 1780, 1899, 1915, 1979, 2007, 2024, 2245, and 2325.

5. The library of claim 1, wherein the library comprises a polynucleotide differentially expressed in a human breast cancer cell, where the polynucleotide comprises a sequence selected from the group consisting of SEQ ID NOS:15, 36, 44, 45, 89, 146, 154, 159, 165, 174, 172, 183, 203, 261, 364, 366, 387, 419, 420, 496, 503, 510, 512, 529, 552, 560, 564, 570, 590, 606, 644, 646, 693, 707, 711, 726, 746, 754, 756, 875, 902, 921, 942, 990, 1095, 1104, 1122, 1131, 1142, 1170, 1184, 1205, 1286, 1289, 1354, 1387, 1435, 1535, 1751, 1764, 1777, 1795, 1860, 1869, 1882, 1890, 1915, 1933, 1934, 1979, 1980, 2007, 2023, 2040, 2059, 2223, 2245, 2300, 2325, 2409, 2462, 2486, 2488, and 2492.

6. The library of claim 1, wherein the library comprises a polynucleotide differentially expressed in a human colon cancer cell, where the polynucleotide comprises a sequence selected from the group consisting of SEQ ID NOS:33, 65, 228, 250, 252, 253, 280, 282, 355, 370, 387, 443, 460, 491, 545, 560, 581, 603, 680, 693, 703, 704, 716, 726, 746, 752, 753, 1095, 1104, 1205, 1241, 1264, 1354, 1387, 1401, 1442, 1514, 1734, 1742, 1780, 1851, 1899, 1915, 1954, 2024, 2066, 2262, and 2325.

7. The library of claim 1, wherein the library comprises a polynucleotide differentially expressed in a human lung cancer cell, where the polynucleotide comprises a sequence selected from the group consisting of SEQ ID NOS: 10, 54, 65, 171, 174, 203, 252, 253, 254, 285, 419, 420, 466, , 491, 525, 526, 552, 571, 574, 590, 693, 700, 726, 742, 746, 861, 990, 922, 1088, 1288, 1355, 1417, 1422, 1444, 1454, 1570, 1597, 1979, 2007, 2024, 2034, 2038, 2126, and 2245.

8. The library of claim 1, wherein the library comprises a polynucleotide differentially expressed in a human cancer cell, where the polynucleotide comprises a sequence selected from the group consisting of SEQ ID NOS:648 and1899.

9. An isolated polynucleotide comprising a nucleotide sequence having at least 90% sequence identity to an identifying sequence of SEQ ID NOS:1-3544, 3546-4510, 4512-4725, 4727-4748, and 4750-5252, or a degenerate variant or fragment thereof.

10. The polynucleotide of claim 9, wherein the polynucleotide comprises a sequence of one of SEQ ID NOS:2503, 2504, 2550, 2555, 2578, 2656, 2667, 2712, 2723, 2728, 2738, 2734, 2754, 2758, 2760, 2832, 2835, 2842, 2843, 2849, 2893, 2933, 2956, 2971, 2981, 3009, 3018, 3019, 3046, 3084, 3190, 3129, 3173, 3226, 3227, 3274, 3290, 3356, 3365, 3377, 3381, 3390, 3391, 3404, 3407, 3408, 3409, 3418, 3419, 3451, 3597, 3600, 3618, 3632, 3635, 3646, 3648, 3657, 3665, 3669, 3670, 3671, 3656, 3680, 3686, 3695, 3696, 3700, 3710, 3736, 3762, 3763, 3774, 3775, 3791, 3804, 3806, 3836, 3895, 3905, 3919, 3920, 3927, 3936, 3951, 3974, 3998, 4036, 4038, 4044, 4056, 4072, 4117, 4119, 4152, 4153, 4154, 4172, 4175, 4159, 4175, 4205, 4216, 4223, 4228, 4238, 4241, 4243, 4251, 4253, 4261, 4263,4278, 4288, 4322, 4330, 4343, 4359, 4363, 4364, 4365, 4373, 4375, 4384, 4385, 4406, 4409, 4431, 4434, 4441, 4442, 4444, 4455, 4469, 4473, 4477, 4482, 4489, 4495, 4496, 4498, 4525, 4535, 4536, 4540, 4560, 4616, 4562, 4586, 4605, 4629, 4653, 4654, 4658, 4659, 4660, 4661, 4664, 4665, 4668, 4684, 4682, 4688, 4689, 4710, 4718, 4733, 4724, 4733, 4746, 4755, 4760, 4710, 4777, 4785, 4792, 4794, 4801, 4807, 4821, 4822, 4847, 4850, 4854, 4856, 4866, 4885, 4900, 4901, 4905, 4914, 4925, 4929, 4931, 4943, 4944, 4959, 5111, 5020, 5041, 5046, 5059, 5083, 5090, 5094, 5102, 5125, 5174, 5197, 5208, 5217, 5237, 5239, 5241, 5243, 5248, and 5252.

11. A recombinant host cell containing the polynucleotide of claim 9.

12. An isolated polypeptide encoded by the polynucleotide of claim 9.

13. An antibody that specifically binds a polypeptide of claim 12.

14. A vector comprising the polynucleotide of claim 9.

15. A polynucleotide comprising the nucleotide sequence of an insert contained in a clone deposited as ATCC accession number xx, xx, xx, xx, xx, xx, xx, xx, or xx.

16. A method of detecting differentially expressed genes correlated with a cancerous state of a mammalian cell, the method comprising the step of:

detecting at least one differentially expressed gene product in a test sample derived from a cell suspected of being cancerous, where the gene product is encoded by a gene corresponding to a sequence of at least one of SEQ ID NOS:10, 15, 33, 36, 44, 45, 54, 65, 89, 146, 154, 159, 165, 171, 172, 174, 183, 203, 228, 250, 252, 253, 254, 261, 280, 282, 285, 355, 364, 366, 370, 387, 419, 420, 443, 460, 466, 491, 496, 503, 510, 512, 525, 526, 529, 545, 552, 560, 564, 570, 571, 574, 581, 590, 603, 606, 644, 646, 648, 680, 693, 700, 703, 704, 707, 711, 716, 726, 742, 746, 752, 753, 754, 756, 861, 875, 902, 921, 922, 942, 990, 1088, 1095, 1104, 1122, 1131, 1142, 1170, 1184, 1205, 1286, 1288, 1289, 1354, 1355, 1387, 1417, 1435, 1444, 1454, 1535, 1570, 1597, 1734, 1742, 1751, 1764, 1777, 1780, 1795, 1860, 1869, 1882, 1890, 1899, 1915, 1933, 1934, 1954, 1979, 1980, 2007, 2023, 2024, 2034, 2040, 2059, 2126, 2223, 2245, 2262, 2300, 2325, 2409, 2486, 2462, 2488, 2492, 1241, 1264, 1401, 1422, 1442, 1514, 1851, 1915, 2007, 2024, 2038, 2066, and 2245;

wherein detection of the differentially expressed gene product is correlated with a cancerous state of the cell from which the test sample was derived.

17. The method of claim 16, wherein said detecting step is by hybridization of the test sample to a reference array, wherein the reference array comprises an identifying sequence of at least one of SEQ ID NOS: 65, 174, 203, 252, 253, 387, 419, 420, 491, 552,

560, 581, 590, 648, 693, 726, 746, 990, 1095, 1124, 1205, 1354, 1387, 1780, 1899; 1915, 1979, 2007, 2024, 2325, and 2245.

18. The method of claim 16, wherein the cell is a breast tissue derived cell, and the differentially expressed gene product is encoded by a gene corresponding to a sequence of at least one of SEQ ID NOS:36, 44, 45, 89, 146, 154, 159, 165, 172, 174, 183, 203, 261, 364, 366, 387, 419, 420, 496, 503, 510, 512, 529, 552, 560,564, 570, 590, 606, 644, 646, 693, 707, 711, 726, 746, 754, 756, 875, 902, 921, 942, 990, 1095, 1104, 1122, 1131, 1142, 1170, 1184, 1205, 1286, 1289, 1354, 1387, 1435, 1535, 1751, 1764, 1777, 1795, 1860, 1869, 1882, 1890, 1915, 1933, 1934, 1979, 1980, 2007, 2023, 2040, 2059, 2223, 2245, 2300, 2325, 2409, 2462, 2486, 2488, and 2492.

19. The method of claim 16, wherein the cell is a colon tissue derived cell, and the differentially expressed gene product is encoded by a gene corresponding to a sequence of at least one of SEQ ID NOS:33, 65, 228, 250, 252, 253, 280, 282, 355, 370, 387, 443, 460, 491, 545, 560, 581, 603, 680, 693, 703, 704, 716, 726, 746, 752, 753, 1095, 1104, 1205, 1241, 1264, 1354, 1387, 1401, 1442, 1514, 1734, 1742, 1780, 1851, 1899, 1915, 1954, 2024, 2066, 2262, and 2325.

20. The method of claim 16, wherein the cell is a lung tissue derived cell, and the differentially expressed gene product is encoded by a gene corresponding to a sequence of at least one of SEQ ID NOS: 10, 54, 65, 171, 174, 203, 252, 253, 254, 285, 419, 420, 466, 491, 525, 526, 552, 571, 574, 590, 693, 700, 726, 742, 746, 861, 922, 990, 1088, 1288, 1355, 1417, 1422, 1444, 1454, 1570, 1597, 1979, 2007, 2024, 2034, 2038, 2126, and 2245.

21. The method of claim 16, wherein the differentially expressed gene product is encoded by a gene corresponding to a sequence of at least one of SEQ ID NOS:648 and 1899.

<223> n = A,T,C or G

<400> 2921
gttactcctc tnanatcagc tacttganga tccctcgatt ngaattcngc acgaggcgat 60
ttatttnaca gagttaaggg gccagtacac ttnatggtat aaaattatct ttntcagggg 120
atgaaggcac aaggagaaaa ttacttgaag cttggagatc ttctctggca agcaatttac 180
aaattctggt gttcttngat ctggctcccn gcccagacaa ccanggagtt nttnatgttc 240
tatcctcatg tgnnannact atacgcaata attngncntn ngccatanag gagggatccg 300
atanntgaca tngctntccn ncanatatac tncncntgna atgnnnctna taatgcatnn 360
nntnnattcc tntctaggnt acncnnantt atatntnntn ggnaactcat ttaacancaa 420
nttcacngca ttcccntggg gttacatata cnctnaagac tatgctgana ctgtgcacca 480
tgnctacatn ngggaattgg atggggtgct tnacggactn ccttgnatgc aagnacttac 540
cagacgtttc canccaanct gacattgntg naatgcatta cncacntggt gntncaantt 600
tactacacct cganaggacc gttcacnggn atttaacctn tcaaanatng ttcnnanggt 660
tacaaggtcc ccaattgtnn gancccttggg gctttgncaa cn 702

<210> 2922
<211> 708
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(708)
<223> n = A,T,C or G

<400> 2922
anaccnttta nnctngttct ttttgcagga tcccatcgat tcgaattcgg cacgaggtat 60
actttgacac tgagaacaaa gagacagtta tatctggaat gggagaatta cacctggaaa 120
tctatgctca gaggctggaa agagagtatg gctgtccttg tatcacagga aagccaaaag 180
ttgcctttcg agagaccatt actgcccctg tcccgtttga ctttacacat aaaaaacaat 240
caggtggtgc aggccagtat ggaaaagtaa taggtgtcct ggagcctctg gacccagagg 300
actacactaa attggaattt tcagatgaaa cattcggatc aaatattcca aagcagtttg 360
tgcctgctgt agaaaagggg tttttagatg cctgcgagaa gggccctctt tctggtcaca 420
agctctctgg gctccggttt gtcctgcaag atggagcaca ccacatggtt gattctaatg 480
aaatctcttt catccgagca ggagaaggtg ctcttaaaca agccttggca aatgcaacat 540
tatgtattct tgaacctatt atggctgtgg aagttgtagc tccaaatgaa tttcagggac 600
aagtaattgc aggaattaac cgacgccatg ggtaatcac tgggcaagat ggagttgagg 660
actattttac actgtatgca gatgtccctc taaatgatat gttgggnt 708

<210> 2923
<211> 715
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(715)
<223> n = A,T,C or G

<400> 2923
gnnnnnttct aatgcnnggc tnttntgcag gatcccatcg attcgctccc attcccggaa 60
ggaggagaca gttactgtct atcccgcaga cgtggtgctc tttgaaggga tcctggggca 120
gaatgaggtg gactatcgcc agaagcaggt ggtcatcctg agccaggata gcttctaccg 180
tgtccttacc tcggagcaga aggccaaagc cctgaagggc cagttcaact ttgaccaccc 240
ggatgccttt gacaatgaac tcattctcaa aacactcaaa gaaatcactg aagggaaaac 300

agtccagatc cccgtgtatg actttgtctc ccattcccag gaggtacgag acctgttcca 360
gatgaagctt tttgtggata cagatgcgga caccccggctc tcacgcagag tattaaggga 420
catcagcgag agaggcaggg atcttgagca gattttatct cagtacatta cgttcgtcaa 480
gcctgccttt gaggaattct gcttgccaac aaagaagtat gctgatgtga tcatccctag 540
aggtgcagat aatctggtgg ccatcaacct catcgtgcag cacatccagg acatcctgaa 600
tggagggccc ttcaaacggc agaccaatgg ctgtctcaac ggctacaccc cttcacgcaa 660
gangcangca tcggagtnca gcagcaggcc gcattgaccc gtcttcatcg gaccc 715

<210> 2924
<211> 724
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(724)
<223> n = A,T,C or G

<400> 2924
gggnctttan atctataggn tacaggctac ttgttctttt tgcaggatcc catccgatgc 60
gcaagtaaga aaacatggcg gctatccttc tctcacatcg aaaaggaaat tttgaacaat 120
catggaaaat ctnggncgtg ctgngaaaac anagaagaga aatgttgcag gaaagattgt 180
ttaanactaa tgaaatacct tttagaacag ctganagaaa ggtttaacng acaaaaanca 240
tctggataaa tnntcttctt atcatgtgaa aactgccttc tttnacntat gtncccagna 300
ccctcaanac agtcagtgng accanacnga nctggncctn tgctttgana actggatgac 360
attcttgntn nattgcctna ggtcagatnn acttgagaat tagttcatcc nnncttcaat 420
ctatcctctt gcagaattnt ttgacatnta cntcagcaat ntttgctnta ncanagnccn 480
atgtaggata tctatgacct nncanngttt gatgantncn tgcnnctgna tnnnncgaga 540
gatntcctaa cnatnncann nnntaanttc tggtantgct caacagattg gaaaaagggg 600
ccaganctgt gnctnaangg ttaaaancnc agganmagta ttttncgtaa acatgnaaan 660
gnttangact gttcatnnnt tgntcctccg aaantgggca cccnttntta ttnattccnc 720
tgcg 724

<210> 2925
<211> 748
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(748)
<223> n = A,T,C or G

<400> 2925
ggtttantt aaatccntnc ncagctactt gttctttttg caggatccca tcgattcgaa 60
ttcggcacga gcggacccat cggagcgtaa cctggatctc cgcaggcctg gcggaggccg 120
gccacctgga ggggcattgc ttggttcgcg tggtancaga ggagcttgag aatgttcgca 180
tcttaccaca tacagttctt tacatggctg attcagaaac tttcattagt ctggaagagt 240
gtcgtggcca taagagagca aggaaaagaa ctagtatgga aacagcactt gcccttgaga 300
agctattccc caaacaatgc caagtccttg ggattgtgac cccaggaatt gtagtgactc 360
caatgggatc angtagcaat cgacctcagg aaatagaaat tggagaatct ggttttgctt 420
tattattccc ttcaaattga aggaataaaa atncaaccct ttcattttat taaggatcca 480
aagaatttaa cattagaaag acatnaactt actgaagtag gtcttttaga taccctgaac 540
ttcgtgtggt cttgnctttg gttataattg ctgtaaggtg ggagccagta attatctgca 600
gcaagtagtc acnctttca gtgatatgaa tatcatcttt ggcttggang ccantngaca 660
acctgncatt actgactttt tgaaaanaac cctctggata ttgatgcctc gggtgtggtt 720